doi:10.1016/j.jmb.2005.09.039 J. Mol. Biol. (2005) 353, 1093–1105

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Crystal Structures of an Intein from the Split *dnaE* Gene of *Synechocystis* sp. PCC6803 Reveal the Catalytic Model Without the Penultimate Histidine and the Mechanism of Zinc Ion Inhibition of Protein Splicing

Ping Sun[1,2]†, Sheng Ye[1,2]†, Sebastien Ferrandon[3], Thomas C. Evans[3] Ming-Qun Xu[3]* and Zihe Rao[1,2]*

[1]*Laboratory of Structural Biology, Tsinghua University Beijing 100084 People's Republic of China*

[2]*National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Science, Beijing 100101 People's Republic of China*

[3]*New England Biolabs Inc. 240 County Road, Ipswich MA 01938, USA*

**Corresponding authors*

The first naturally occurring split intein was found in the *dnaE* gene of *Synechocystis* sp. PCC6803 and belongs to a subclass of inteins without a penultimate histidine residue. We describe two high-resolution crystal structures, one derived from an excised Ssp DnaE intein and the second from a splicing-deficient precursor protein. The X-ray structures indicate that His147 in the conserved block F activates the side-chain N^{δ} atom of the intein C-terminal Asn159, leading to a nucleophilic attack on the peptide bond carbonyl carbon atom at the C-terminal splice site. In this process, Arg73 appears to stabilize the transition state by interacting with the carbonyl oxygen atom of the scissile bond. Arg73 also seems to substitute for the conserved penultimate histidine residue in the formation of an oxyanion hole, as previously identified in other inteins. The finding that the precursor structure contains a zinc ion chelating the highly conserved Cys160 and Asp140 reveals the structural basis of Zn^{2+}-mediated inhibition of protein splicing. Furthermore, it is of interest to observe that the carbonyl carbon atom of Asn159 and N^{η} of Arg73 are 2.6 Å apart in the free intein structure and 10.6 Å apart in the precursor structure. The orientation change of the aromatic ring of Tyr-1 following the initial acyl shift may be a key switching event contributing to the alignment of Arg73 and the C-terminal scissile bond, and may explain the sequential reaction property of the Ssp DnaE intein.

Keywords: Ssp DnaE; split intein; protein splicing; crystal structure; zinc inhibition

Introduction

Protein splicing is an extraordinary post-translation rearrangement that involves the precise removal of an internal protein sequence, termed an intein, from a precursor protein with the concomitant ligation of the flanking polypeptide sequences, termed exteins.[1–3] Since its first report in 1990, more than 200 inteins, ranging from 128 to 1650 amino acid residues, have been identified and deposited into the New England Biolabs Intein Database.[4]

The protein self-splicing pathway was deciphered by the identification of key intermediates and catalytic residues that are inherent in a set of highly conserved sequence motifs shared by all known inteins. The splicing reaction consists of a series of concerted nucleophilic replacements.[5–7] The N-terminal intein segment (in the range of 100–150 residues) possesses the conserved blocks A

† P.S. and S.Y. made equal contributions to this work.

Abbreviations used: Nsp, *Nostoc* sp. PCC 7120; Ssp.6803, *Synechocystis* sp. PCC 6803; Tel, *Thermosynecococcus elongatus* BP-1; Ter, *Trichodesmium erythraeum*; Ssp.6301, *Synechococcus elongatus* PCC 6301; Ssp.7002, *Synechococcus* sp PCC 7002; Sel, *Synechococcus elongatus* PCC7942; Aha, *Aphanothece halophytica*; Oli, *Oscillatoria limnetica*; Cwa, *Crocosphaera watsonii* WH8501; Ava, *Anabaena variabilis* ATCC29413; Aov, *Aphanizomenon ovalisporum*; Tvu, *Thermosynecococcus vulcanus*; Npu, *Nostoc punctiforme*.

E-mail addresses of the corresponding authors: xum@neb.com; raozh@xtal.tsinghua.edu.cn

and B, whereas the C-terminal segment of approximately 35–50 residues contains blocks F and G (Figure 1(a)).[8] A typical intein initiates the protein splicing reaction by an acyl shift involving the side-chain of a cysteine or serine residue in block A at the N-terminal splice junction. The first C-extein residue following the scissile bond at the C-terminal splice site is invariably cysteine, serine, or threonine, and the sulfhydryl or hydroxyl group on their side-chain attacks the thioester (or ester) bond at the N-terminal splice junction, resulting in a branched intermediate. Next, scission of the peptide bond at the C-terminal splice junction is coupled with side-chain cyclization of the highly conserved asparagine residue. Finally, the occurrence of a spontaneous S–N or O–N acyl rearrangement forms a peptide bond between the newly ligated extein sequences. Most inteins possess an HN/C (S or T) motif in block G flanking the C-terminal splice site. The penultimate histidine residue appears to play an important role in splicing by assisting cyclization of asparagine.[9–13]

The three-dimensional structures of the vacuolar ATPase subunit intein from *Saccharomyces cerevisiae* (Sce VMA intein),[14–19] the archaeal PI-PfuI intein from *Pyrococcus furiosus* (PI PfuI intein),[20] the mini-intein of the GyrA protein from *Mycobacterium xenopi* (Mxe GyrA intein)[21] and the mini-intein of the DnaB from *Synechocystis* sp. PCC6803 (Ssp DnaB intein)[9] reveal that inteins possess a horseshoe-like β-strand scaffold termed the Hedgehog/Intein (Hint) module.[22] These studies further indicate that the conserved block B, F and G residues are responsible for both splicing and the N and C-terminal cleavage activity. However, little structural information is available for the non-canonical inteins, which may have acquired an alternative splicing mechanism during evolution.

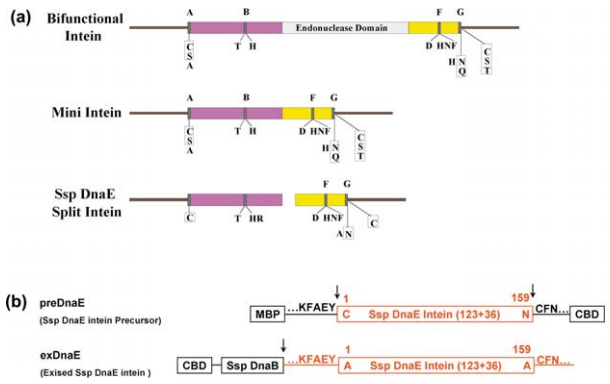

Figure 1. (a) Diagram of conserved intein motifs of bifunctional inteins, mini inteins and the Ssp DnaE split intein. Blocks A and B (black) in the N-terminal subdomain (magenta) and blocks F and G (black) in the C-terminal subdomain (yellow) are shared by the splicing domains and the endonuclease domain is shown in grey. Residues involved in nucleophilic attack (letters in a box), as well as other highly conserved amino acids are indicated below the block diagram. (b) A representation of the Ssp DnaE intein fusions. The exDnaE fusion protein consists of maltose-binding protein (MBP), the full-length wild-type Ssp DnaE intein (residues 1–159, which include 123 amino acid residues from the N terminus and 36 amino acid residues from the C terminus) with five native extein residues at its N terminus and three native residues at its C terminus, and the CBD. The resulting protein exDnaE is splicing functional. Black arrows indicate the splicing sites of Ssp DnaE intein. preDnaE consists of CBD, Ssp DnaB intein and the full-length Ssp DnaE intein with C1A and N159A mutations (residue 1–159) along with five native extein residues at its N terminus and three native residues at its C terminus. The black arrow shows the cleavage site of the modified Ssp DnaB intein.[41] The intein proteins after purification are indicated in red.

The Ssp DnaE intein is a naturally occurring split intein from the split *dnaE* gene of *Synechocystis* sp. PCC6803. The *dnaE* gene encoding the catalytic subunit of the bacterial DNA polymerase III is separated into two coding regions. The first region consists of the 774 N-terminal amino acid residues of the DnaE protein, followed by the 123 N-terminal amino acid residues of the Ssp DnaE intein sequence. The second region consists of the 36 C-terminal amino acid residues of the Ssp DnaE intein followed by the 423 C-terminal amino acids of the DnaE polymerase. These two split genes are located 745 kb apart on the genome, and are found on opposite strands of the DNA.[23] Formation of the full-length host protein is speculated to occur by the protein splicing activity of the N and C-terminal Ssp DnaE intein segments. When these two intein halves are co-expressed in *Escherichia coli*, they are capable of *trans*-splicing to produce intact DNA polymerase III. Mechanistic studies of protein self-splicing have been facilitated by the use of the Ssp DnaE intein, since the two reactive components can be separately expressed, purified and reassembled for *in vitro* analysis. The Ssp DnaE intein possesses *trans* and *cis*-splicing activities both *in vivo* and *in vitro*.[24] Furthermore, the Ssp DnaE intein belongs to a subclass of *trans*-splicing inteins found in cyanobacteria, and these non-canonical inteins lack the highly conserved penultimate histidine.

In order to explore the splicing mechanism of the Ssp DnaE intein, we have determined the crystal structures of an excised Ssp DnaE intein termed exDnaE, and a splicing-deficient precursor protein termed preDnaE (Figure 1(b)). The failure to co-crystallize the N and C termini of Ssp DnaE split intein in our initial experiments led to the construction of a *cis*-splicing system for structural and mechanistic studies. Furthermore, zinc is the only confirmed reversible inhibitor of inteins to date. Biochemical studies have shown that zinc can inhibit both *trans*-splicing and N-terminal cleavage activities of the Ssp DnaE intein.[10,25–27] The structure of the precursor protein with a bound zinc ion reveals the mechanism of zinc ion inhibition. Finally, a unique characteristic of the Ssp DnaE intein lies in the fact that N-terminal cleavage is a prerequisite for C-terminal cleavage.[26,28] The structure of the Ssp DnaE intein provides a structural basis for the elucidation of this sequential catalytic mode. A comparison between the pre and

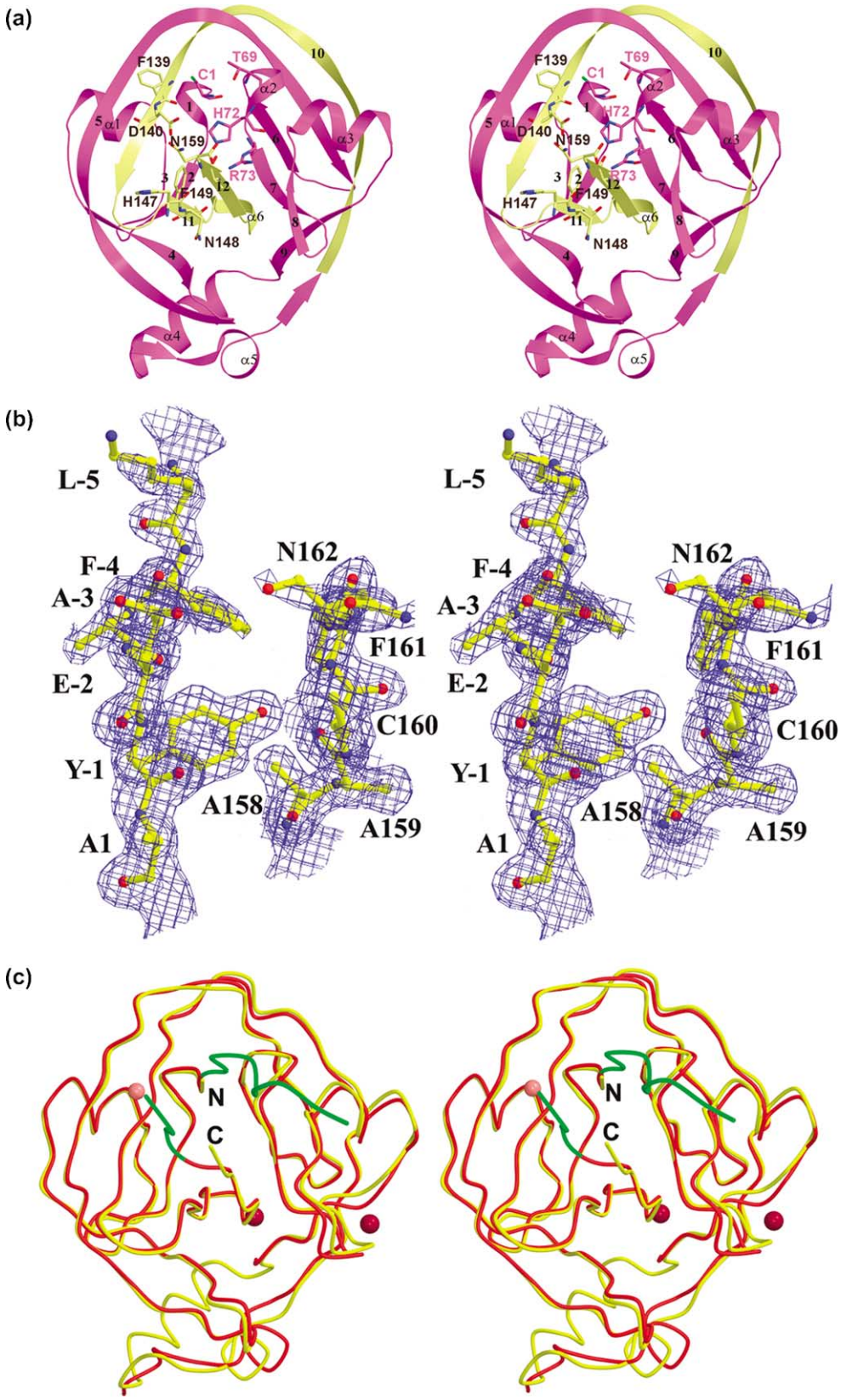

Figure 2 (*legend next page*)

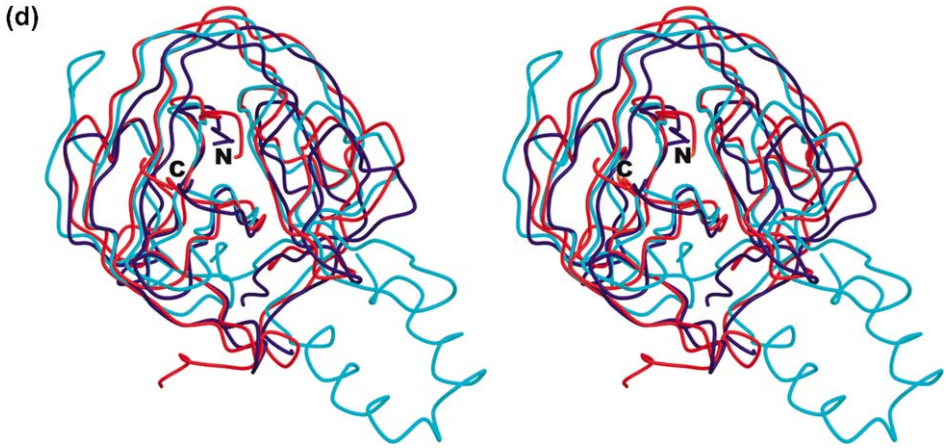

Figure 2. (a) Stereo view of the Ssp DnaE-free intein structure (exDnaE). Ribbon drawing of the C^{α} backbone and the catalytic residues of the Ssp DnaE intein. The α-helices and β-strands are numbered. The N-terminal intein sequence is shown in magenta and the C-terminal intein sequence is shown in yellow. Important residues involved in catalysis are shown in ball-and-stick representation and labeled. (b) Stereo view of the electron density associated with the N and C-terminal additional extein residues present in preDnaE. A $2|F_{obs}| - |F_{calc}|$ map is shown contoured at 1.0σ covering the extein residues with a radius of 2.0 Å. Tyr-1 through Lys-5 are associated with the N-terminal extein, whereas Cys160 through Asn162 are associated with the C-terminal extein. (c) Stereo view of the superposition of the backbone of the preDnaE structure (red) and exDnaE structure (yellow). The N and C-exteins are colored green. The zinc ion and the two calcium ions in the preDnaE of Ssp DnaE intein are shown as flesh-tints sphere and red-brown spheres, respectively. The N and C-terminals are labeled. (d) Stereo view of the superposition of the backbone of the preDnaE structure (red), Ssp DnaB intein structure (dark blue) and Mxe GyrA intein structure (cyan). The N and C-terminals are labeled.

post-splicing intein structures also provides important insights into the general protein splicing mechanism.

Results and Discussion

Structure of the Ssp DnaE intein

The Ssp DnaE intein is composed of an N-terminal segment of 123 residues and a C-terminal segment of 36 residues. The N and C-terminal segments of the Ssp DnaE intein could not be co-crystallized and were instead expressed together as a *cis*-splicing system.

The exDnaE-free intein contains two molecules in the asymmetric unit (chains A and B) that are essentially identical (rmsd 0.8 Å over 159 C^{α} atoms). The lack of shape complementarity and small buried surface area at the dimer interfaces suggest that this interaction is unlikely to be physiologically relevant. For the purposes of discussion, structural analysis refers to chain A of exDnaE (Figure 2(a)).

The preDnaE has one molecule per asymmetric unit. The five native N-extein residues (Lys-5, Phe-4, Ala-3, Glu-2 and Tyr-1) and three native C-extein residues (Cys160, Phe161 and Asn162) are well defined in electron density maps (Figure 2(b)). The preDnaE structure contains two calcium ions (Figure 2(c)), as a likely result of the 200 mM $CaCl_2$ added during crystallization. In addition, an ordered zinc ion is observed. Detailed functional analysis of zinc ion binding will be discussed in a later section.

Both exDnaE and preDnaE structures are compact and composed primarily of β-strands. Superposition of these two structures shows an overall structural similarity with an rmsd difference of 1.1 Å for equivalent C^{α} atoms (Figure 2(c)). The N and C termini are tightly associated, with numerous hydrogen-bonding contacts and hydrophobic packing interactions between them. The two splice junctions are located at the center of the horseshoe-like structure at the ends of two adjacent antiparallel β-strands. In exDnaE, 14 β-strands and six α-helices comprise the whole structure, termed the Hint domain.[22] However, the preDnaE structure lacks the α4 helix, consistent with other mini-intein structures.[9,16] The α4 helix is probably stabilized by crystal packing in the unit cell of exDnaE. It is not clear whether the α4 helix represents the natural conformation under conditions favorable for the unique high affinity between the N and C domains in Ssp DnaE intein or an artefact of crystallization.

Superposition of four other available intein structures indicates that they share overall structural similarity. The rmsd for the C^{α} atoms of the Hint domain is about 1.0 Å between preDnaE and Ssp DnaB inteins,[9] 1.0 Å between preDnaE and Mxe GyrA inteins (Figure 2(d)),[21] 1.6 Å between preDnaE and Sce VMA inteins,[14–16,29] and 1.4 Å between preDnaE and *P. furiosus* Pol inteins.[20]

Structure of the N-terminal catalytic site

Through superposition of the N-terminal subdomain of preDnaE and exDnaE structures (Figure 3(a)), the S-H bond of the Cys1 was modeled into the precursor structure and found to be polarizable by hydrogen bonding with Arg50

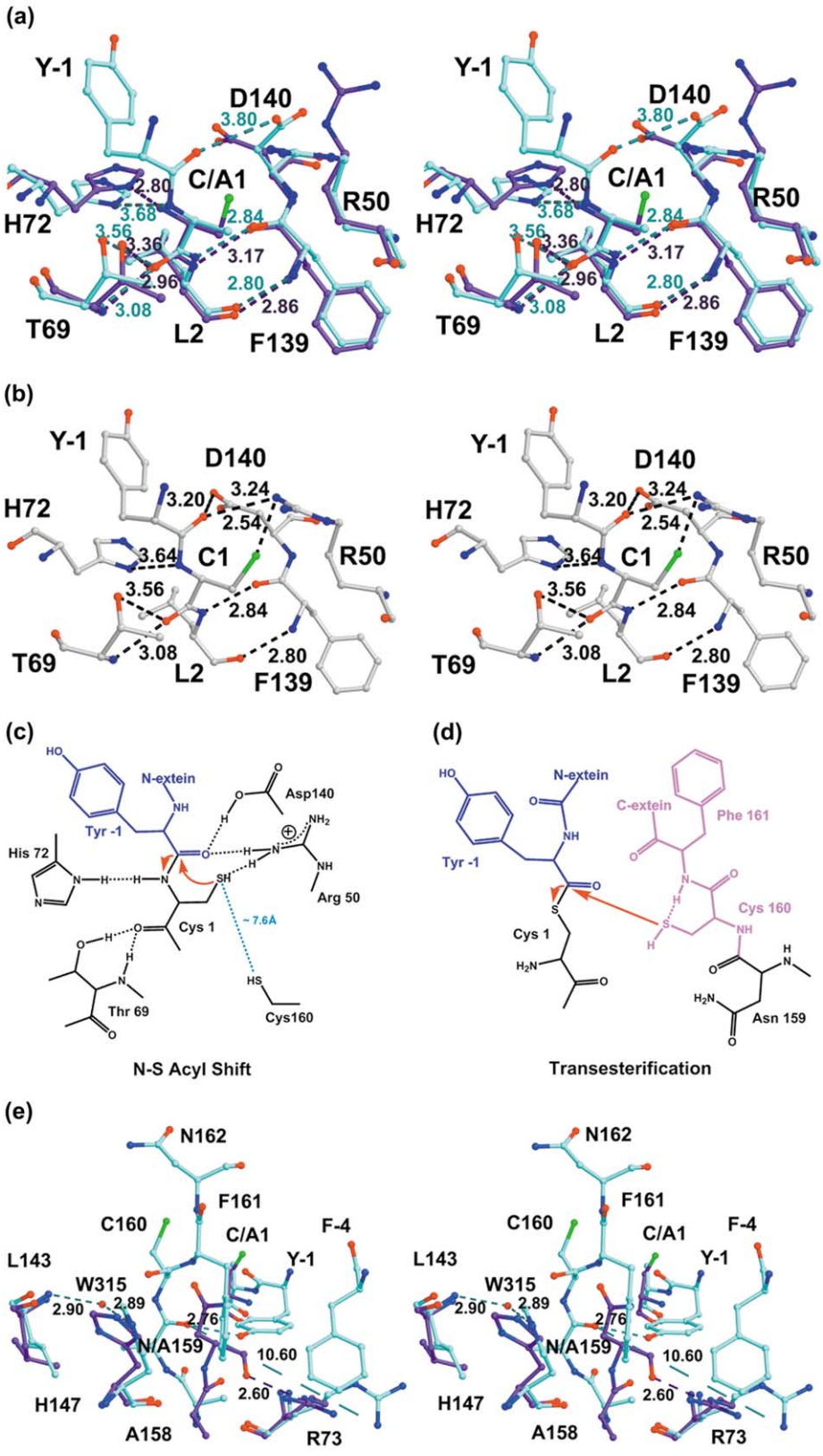

Figure 3 (*legend next page*)

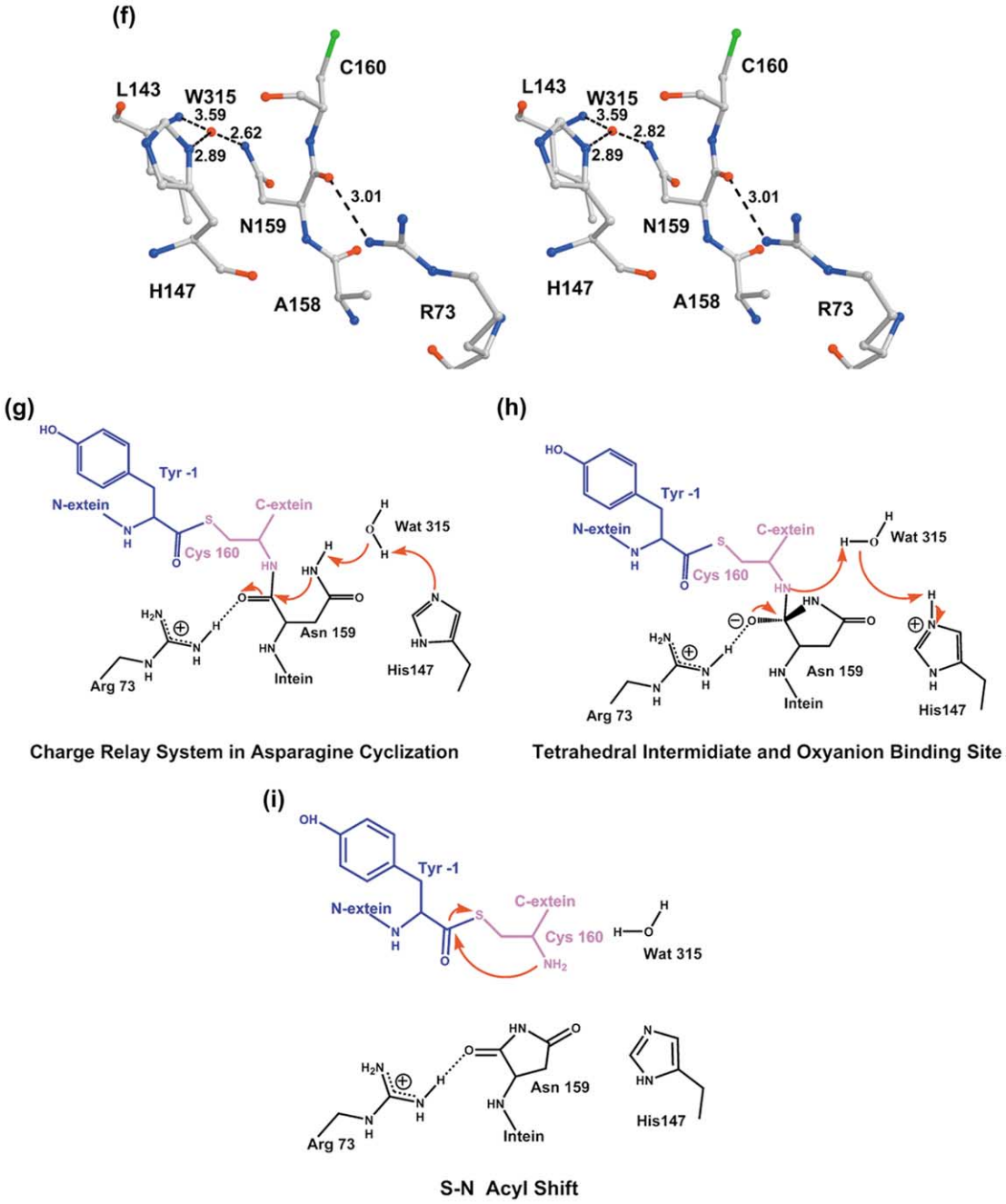

Figure 3. (a) Close-up stereo view of the superposition of the N-terminal subdomain of preDnaE (cyan) and exDnaE (purple). (b) Stereo view of the modeled N-terminal catalytic module of Ssp DnaE intein. (e) Close-up stereo view of the superposition of the C-terminal subdomain of preDnaE (cyan) and exDnaE (purple). (f) Stereo view of the modeled C-terminal catalytic module of Ssp DnaE intein. Residues are shown in ball-and-stick representations. The broken lines indicate hydrogen bonds, and bond distances are labeled. (c), (d), (g), (h) and (i) A chemical mechanism proposed for splicing the Ssp DnaE intein. The red arrows indicate the routes of nucleophilic attacks in the splicing pathway. Broken lines indicate hydrogen bonds. The tetrahedral intermediate formed by an N-S acyl rearrangement at Cys1 is not shown.

(Figure 3(b)). The modeled nitrogen atom of Cys1 is at a distance of 3.6 Å from the His72 N^{δ} atom, while the modeled oxygen atom is 3.6 Å from the Thr69 hydroxyl oxygen atom and 3.1 Å from Thr69 amide nitrogen atom. The side-chain of Asp140 is hydrogen-bonded with the carbonyl oxygen atom of Tyr-1. The amide nitrogen atom of Ile2 is 2.8 Å from the carbonyl oxygen atom of Phe139. The arrangement of this active site is capable of facilitating the initial step in the splicing pathway, a nucleophilic attack by the thiol of Cys1 on the carbonyl carbon atom of the peptide bond between Cys1 and Tyr-1 to form a thioester intermediate (Figure 3(c) and (d)). Contrary to the *Mxe* GyrA intein structural data,[21,30] the scissile bond at the N-terminal splice junction is in the ordinary *trans* conformation. The interactions of Cys1 with Arg50, Thr69, His72 and Ile2 with Phe139 within the active

site may contribute to the retention of the strained conformation of this residue. Replacement of Thr69, His72 or Phe139 with alanine caused the failure of both N-terminal cleavage and splicing.[10]

C-terminal catalytic site reveals the catalytic pattern of asparagine cyclization

In canonical inteins, the last residue is asparagine and the penultimate residue is histidine.[5,7,31] Mutagenesis studies have shown that the penultimate residue plays an important role in the processing of asparagine cyclization.[9–13] However, non-canonical inteins are able to undergo efficient protein splicing with a penultimate residue other than histidine. Replacement of the penultimate residue with a histidine in non-canonical inteins can either reduce, enhance, or have no measurable effect on the extent of splicing.[23,32] These investigations imply an alternative mechanism involved in the splicing steps. Here, we report the first intein structure with a naturally occurring alanine residue in the penultimate position.

Replacement of the penultimate alanine residue in the Ssp DnaE intein with histidine has no measurable effect on the extent of splicing.[25] Superposition of preDnaE structure with the Ssp DnaB intein structure shows that, although these two structures share similarities in the C-terminal catalytic center, a hydrophobic area in the Ssp DnaE intein formed by the native N-extein residue Phe-4 and C-extein residue Phe161 forces the imidazole ring of the modeled penultimate histidine away from the carbonyl oxygen atom of Asn159. As a consequence, the imidazole ring in the A158H mutant is not ideally placed to assist in cyclization of Asn159 in the Ssp DnaE intein.

The hydrogen bond between the carbonyl oxygen atom of Asn159 and N^{η} of Arg73 is about 2.6 Å in the exDnaE-free intein structure. In constrast, the corresponding distance is about 10.6 Å in the preDnaE intein structure (Figure 3(e)). Therefore, we propose that Arg73 could stabilize the negatively charged carbonyl oxygen atom of Asn159 *via* hydrogen bonds upon tetrahedral intermediate formation (Figure 3(f)).

The charge relay system identified in the Ssp DnaB intein structure is present also in the Ssp DnaE intein (Figure 3(f) and (g)). A highly ordered water molecule, Wat315, occupies the position equivalent to that of Wat30 in the Ssp DnaB intein. The N^{δ} atom of His147 accepts a proton from N^{δ} of Asn159. In turn, the deprotonated N^{δ} atom initiates a nucleophilic attack on the carbonyl carbon atom, resulting in a break of the main chain. His147 and Asn159 form hydrogen bonds with their neighboring residues, which contribute to their conformational stabilities. The highly ordered water molecule (Wat315) is also stabilized by a hydrogen bond with Leu143. Upon tetrahedral intermediate formation, the negatively charged carbonyl oxygen atom of Asn159 is stabilized by a hydrogen bond with N^{η} of Arg73 (Figure 3(h)). Collapse of the tetrahedral intermediate is facilitated by the transfer of a proton to the leaving amide group from Wat315, which in turn accepts a proton from the imidazolium side-chain of His147. A spontaneous O-N acyl shift results in a stable peptide bond between Tyr-1 and Cys160 (Figure 3(i)). In this process, it is Arg73 that acts as the substitute for the missing penultimate histidine. In addition, inteins with a penultimate histidine residue in close interaction with the carbonyl oxygen atom of the adjacent C-terminal asparagine residue might also contribute to the proper orientation of the asparagine's side-chain.

Sequence alignment of all 14 split DnaE inteins lacking the conserved penultimate histidine residue[4,33] shows that the residue equivalent to Arg73 in block B in Ssp DnaE intein is either arginine (six out of 14 species), lysine (seven out of 14 species) or glutamine (one out of 14 species) (Table 1). The three residue types (arginine, lysine, and glutamine) all possess a flexible side-chain and could form a hydrogen bond with the carbonyl oxygen atom of Asn159, which stabilizes the negatively charged carbonyl oxygen atom of Asn159 upon tetrahedral intermediate formation.

Table 1. Sequence alignment of split DnaE inteins

	Block A	Block B	Block F	Block G
Nsp DnaE	CLSYDTEVLTVEY N13	GSIIKATKDHKFMT N76	NVYDIGVRRDHNFF C27	NGLIASNC C37
Ssp 6803 DnaE	CLSFGTEILTVEY N13	GSVIRATSDHRFLT N76	RIFDIGLPQDHNFL C27	NGAIAANC C37
Tel DnaE	CLSGETAVMTVEY N13	GSTICATPDHRFMT N76	AVYDIGLAADHNFV C26	NGAIAANC C36
Ter DnaE-3	CLTYETEIMTVEY N13	GTVIRATPEHKFMT N76	NVYDIGVTKDHNFV C27	NGLIASNC C37
Ssp.6301 DnaE	CLAADTEVLTVEY N13	GRIIRATADHRFMT N76	PVYDLGVATVHNFV C27	NGLVASNC C37
Ssp.7002 DnaE	CLAGGTPVVTVEY N13	GQMIRATPDHRFLT N76	PTYDIGLSQDHNFL C27	QGLIAANC C37
Sel DnaE	CLAADTEVLTVEY N13	GRIIRATADHRFMT N76	PVYDLGVATVHNFV C27	NGLVASNC C37
Aha DnaE	CLSYDTEIWTVEY N13	GRKIRATKDHKMMT N76	NVYDVCVETDHNFV C27	NGCVASNC C37
Oli DnaE	CLSYNTEVLTVEY N13	GSVIRATKDHQFMT N76	NVYDIGVEKDHNFL C27	SGEIASNC C37
Cwa DnaE	CLSYDTEILTVEY N13	GSKIKATKDHKFMT N76	KVYDIGVEKDHNFL C27	NGSIASNC C37
Ava DnaE	CLSYDTEVLTVEY N13	GSIIKATKDHKFMT N76	NVYDIGVGRDHNFF C27	NGLIASNC C37
Aov DnaE	CLSADTEILTVEY N13	GSIIRATKDHKFMT N76	NVYDIGVEHHHNFA C27	NGLIASNC C37
Tvu DnaE	CLSGETAVMTVEY N13	GSTICATPDHRFMT N76	AVYDIGLAGDHNFL C26	NGAIAANC C36
Npu DnaE	CLSYETEILTVEY N13	GSLIRATKDHKFMT N76	NVYDIGVERDHNFA C27	NGFIASNC C37
Total:14		Arg(R)6 Lys(K)7 Gln(Q)1		Ala(A)4 Ser(S) 10

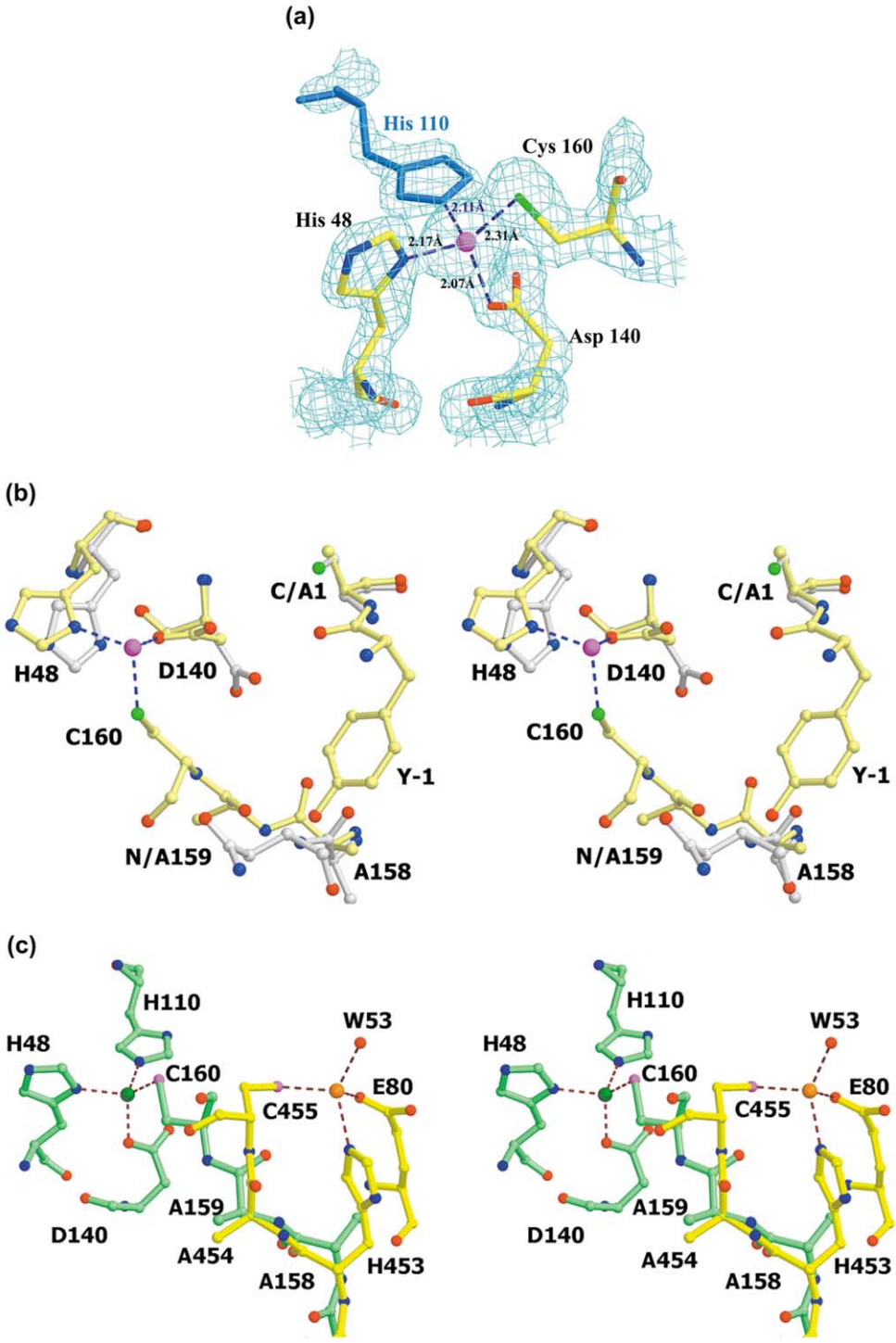

Figure 4. The zinc coordination module. (a) Close-up view of the zinc coordination at the C-terminal splicing junction of the Ssp DnaE split intein. Four coordinating residues, His48, Asp140, His110 (from a neighbouring molecule) and the highly conserved residue Cys160, are shown in ball-and-stick representation. The zinc ion is shown as a purple sphere. The distances between zinc ion and coordinating ligands are shown. The $2|F_{obs}|-|F_{calc}|$ map is contoured at 2.0σ with a cover radius of 2.0 Å. (b) Stereo view of the superposition of residues involved in zinc coordination from preDnaE (yellow) with exDnaE (white) structures. The purple sphere represents the zinc ion in the preDnaE structure. (c) Stereo view of the superposition of residues involved in zinc coordination from the preDnaE of the Ssp DnaE intein (green) and the PI-SceI intein (PDB number 1EF0, molecule B; yellow). The dark green sphere and orange sphere represent the zinc ions in the preDnaE structure and PI-SceI structure, respectively.

Therefore, the mechanism by which Arg73 contributes to the asparagine cyclization process appears to be conserved in all split DnaE inteins lacking the conserved penultimate histidine ressidue.

The distance between S^{γ} of the modeled side-chain of Cys1 and the sulfur group of Cys160 in the precursor of Ssp DnaE intein structure is 7.6 Å, compared with 8.5 Å for the Ssp DnaB intein precursor (containing the C1A and N154A mutations) and 9 Å for the Sce VMA intein precursor (containing the C1A and N454A mutations).[9,16] These distances are each longer than the 3.6 Å between N and C termini of the Sce VMA intein precursor, which was trapped by a different set of amino acid substitutions (Cys1, Asn454, and Cys455 were all substituted with serine).[17] A conformational change would therefore have to occur to traverse this relatively large gap after the initial acyl shift step.

Mechanism of zinc ion inhibition of protein splicing

Zinc ion is the only confirmed reversible inhibitor of inteins to date. A zinc ion was found in the structure of preDnaE. Zinc is coordinated in a T_4 (tetrahedral) geometry by four residues, His48, Asp140, His110 (from a neighboring molecule) and the highly conserved residue Cys160 (Figure 4(a)).[34] The metal–donor atom target distances are all within accepted constraints. The observed zinc binding is consistent with the hypothesis proposed by Nichols *et al*., that a 1:1 ratio of metal ion to intein complex is required for maximum inhibition of the splicing reaction.[26] The zinc ion is located on the surface in the interface between two neighboring molecules. Although zinc is supposedly helpful for the packing of crystals, since crystals grown in the absence of zinc ion diffract poorly, we would prefer to interpret the zinc binding from a functional perspective.

Superposition of preDnaE and exDnaE structures reveals different orientations of the conserved residue Asp140 (Figure 4(b)). The distance between the side-chain oxygen atom and the carbonyl oxygen atom of Tyr-1 is 3.8 Å, whereas the equivalent distance in the modeled Ssp DnaE intein precursor is about 3.2 Å. We conclude that Asp140 could contribute to the stabilization of the tetrahedral intermediate. Zinc binding blocks the function of Asp140, and in turn causes the inhibition by zinc of N-terminal cleavage with either no DTT or a low concentration of DTT added.[10,26,27]

The most critical feature of zinc binding is the involvement of Cys160. It has been reported that the Cys residue at the C-terminal splice junction of the *Mycobacterium tuberculosis* RecA intein has a low apparent pK_a, which would facilitate the occurrence of *trans*-esterification by attacking the thioester bond involving Cys1.[35] Cys160 of the Ssp DnaE intein appears to play an important role in driving the first splicing step and shifting the equilibrium of amide and thioester. Therefore, inactivation of Cys160 by zinc chelation affects the acyl rearrangement involving Cys1, thereby shifting the equilibrium of amide and thioester to favor amide formation. Secondly, chelation of a zinc ion will neutralize the nucleophile present at the C-terminal splice junction, negating the nucleophilic attack at the thioester bond formed at the N-terminal splice junction, and ultimately leading to a failure of splicing. Substitution of Cys160 with any residue including, Thr and Ser, could alter the conformation required for chelating zinc.[10,26]

Zinc inhibition of different types of inteins will vary according to their chelating modules. The phenomenon of zinc ion binding was previously discovered in the crystal structures of the PI-Pful intein from *P. furiosus*,[20] and the PI-SceI intein from *Saccharomyces cerevisiae*.[16] In molecule B of the PI-SceI intein, the zinc ion binding site includes the penultimate histidine residue and the first C-extein cysteine residue that participate in the third and second splicing steps, respectively. In molecule A of the PI-SceI intein, however, the coordinating ligands also include His442 in block F, which is the key residue in proton transfer during asparagine cyclization. Superposition of the PI-SceI intein[16] and preDnaE shows different zinc ion chelating models (Figure 4(c)). The positions occupied by the zinc ions are not identical. Cys+1 (Cys160 in the Ssp DnaE intein and Cys455 in the PI-SceI intein) is the only common chelating residue. Furthermore, mutagenesis studies have revealed that an Ala158His substitution neither increases the Ssp DnaE splicing efficiency nor increases the activity of zinc ion inhibition.[10]

Native extein residues contribute to the sequential catalytic property of the Ssp DnaE intein

The interaction of the two intein halves under natural or artificial conditions is at least partly, or probably entirely, dominated by the association of the intein fragments and not by the extein residues.[24,36] Evans and co-workers showed that replacement of the four native C-extein residues in the Ssp DnaE intein (^{160}Cys-Phe-Asn-Lys) with either ^{160}Cys-Thr-Gly-Met or ^{160}Cys-Phe-Thr-Gly results in retained N-terminal cleavage activity but loss of splicing activity. The authors hypothesized that at least two proximal native N-extein and three proximal native C-extein residues are essential for the Ssp DnaE intein splicing activity in *cis*-splicing systems.[24] The Ssp DnaE split inteins in protein engineering applications retain five native N-extein and three native C-extein residues to maintain their high level of splicing activity.[24,37–39] It has been reported that extein residues can alter the splicing or cleavage activity of a number of inteins. However, the role of exteins has not been fully understood. Furthermore, the effect of extein residues on splicing is of interest for further optimization of protein engineering systems using

the Ssp DnaE split inteins. In preDnaE, the five native N-extein and three native C-extein residues are clearly present (Figure 2(b)). Three pairs of hydrogen bonds were found to conjoin the two adjacent splicing junctions. The hydroxyl group of the N-terminal Tyr-1 residue is within hydrogen-bonding distance (around 2.8 Å) of the Asn159 carbonyl oxygen atom (Figure 3(e)). The first C-extein residue, Cys160, maintains its conformation *via* a hydrogen bond with His147 in block F. In addition, the side-chain of native Asn162 is stabilized by a hydrogen-bond interaction with Gln46. Substitution of Asn162 with Thr, bearing a shorter side-chain, or the more flexible Gly, results in failure to establish a proper interaction between the N and C-splicing junctions.

In contrast with other identified inteins, the Ssp DnaE split intein has a unique property, in that N-terminal cleavage is a prerequisite for C-terminal cleavage.[27] It is therefore of interest to investigate the possible reasons for this sequential catalytic mode. Superposition of preDnaE with exDnaE and the Ssp DnaB intein structures suggests the relative mobility of the Ssp DnaE intein C terminus. We thus propose that the proximate N-extein residue Tyr-1 plays an important role in the sequential reactive property of the Ssp DnaE intein. The hydrogen bond between the Asn159 carbonyl oxygen atom and N^{η} of Arg73 measures about 2.6 Å in exDnaE, whereas the corresponding distance in preDnaE is about 10.6 Å. The distance is reduced greatly after the splicing event occurs. Detailed comparison between preDnaE and exDnaE structures in the C-terminal catalytic center reveals that a change in orientation of the aromatic ring of Tyr-1 might be the key switch (Figure 3(e)). Prior to N-terminal cleavage, the aromatic ring is located between Ala158 and Arg73, and blocks the formation of a hydrogen bond between these residues. Following the first N-S acyl shift and rearrangement of several hydrogen bonds, the orientation of the Tyr-1 side-chain might be changed slightly. This reorientation of Tyr-1 would create enough space for the side-chain of Arg73 to be properly placed to assist in cyclization of Asn159 (Figure 3(f)). Therefore, without the N-terminal cleavage as a prerequisite, the side-chain of the functional residue Arg73 could not be directed to the Asn159 carbonyl oxygen atom, resulting in the failure of asparagine cyclization.

This investigation describes the splicing mechanism for a non-canonical intein lacking the conserved penultimate histidine residue. In the Ssp DnaE split inteins, Arg73 takes the place of the typical penultimate histidine residue and contributes to stabilization of the oxyanion in the tetrahedral intermediate. This observation adds to our understanding of the diversity of intein biochemistry. Previously, the conserved histidine residue in block B has been found to contribute to the scission of the N-terminal splice junction. Significantly, this study reveals the role of Arg73 adjacent to this histidine residue in block B in C-terminal cleavage. This new finding may provide insights into the evolution of the catalytic activities of inteins. As the DnaE split inteins may have evolved from continuous inteins with typical intein motifs, the modification of the asparagine cyclization mechanism could be the result of evolutionary processes. The selective pressure for efficient expression of the DNA polymerase yields an intein that maintains its splicing activity by coupling the reassociation of the two segments with the key residue located at the N-intein. Although this mechanism seems to be plausible in other DnaE split inteins found so far, the splicing mechanisms in all non-canonical inteins remain to be elucidated. Hence, there will be continued interest in inteins due to their useful applications in protein engineering, leading to further understanding of auto-catalysis from the view of structure-function relationships.

Materials and Methods

Protein expression and purification

The DNA sequence encoding the full-length Ssp DnaE intein (residue 1–159, which includes 123 amino acid residues from the N terminus and 36 amino acid residues from the C terminus) with C1A and N159A mutations along with its five native extein residues and three native C-N-extein residues was transferred from pMEB3[24] into the pTWIN vector,[40] yielding preDnaE. Because the key residues involved in splicing steps (C1 and N159) were mutated to alanine, the resulting Ssp DnaE intein is splicing-deficient. The construct exDnaE expresses a fusion protein composed of maltose-binding protein (MBP), the full-length wild-type Ssp DnaE intein (residue 1–159) with five native extein residues at its N terminus and three native residues at its C terminus, and the CBD [24] (Figure 1(b)). The free DnaE intein is generated *via in vivo* splicing during expression from the exDnaE construct.

The preDnaE fusion protein was overexpressed in *E. coli* strain ER2566 with 100 μM $ZnCl_2$ added to the culture medium. The selenomethionyl exDnaE derivative (Se-exDnaE) was produced by expression in the methionine-deficient *E. coli* strain B834 (DE3).

Purification of preDnaE was performed by the pH and temperature-induced intein-mediated cleavage method.[41] The CBD-Ssp DnaB intein-Ssp DnaE intein fusion protein was purified at 277 K using chitin resin (New England Biolabs, Inc.) and cleavage was induced by quickly washing the column with three column volumes of buffer (40 mM sodium phosphate (pH 6.5), 200 mM NaCl, 100 μM $ZnCl_2$) and incubating overnight at 291 K. Fractions were collected by eluting with 10 mM Tris HCl (pH 7.5), 100 μM $ZnCl_2$. The solution was applied to a Resource Q column pre-equilibrated with 10 mM Tris-HCl (pH 7.5), 100 μM $ZnCl_2$. The preDnaE protein was eluted with a linear gradient of NaCl. The peak containing preDnaE was further applied to a Superdex 75 column pre-equilibrated with 10 mM Tris–HCl (pH 7.5), 150 mM NaCl, 100 μM $ZnCl_2$ and eluted with the same buffer.

Se-exDnaE protein was purified using a Q-Sepharose (Fast Flow) column pre-equilibrated with 20 mM Tris-HCl (pH 7.5). The Se-exDnaE protein was eluted with a linear gradient of NaCl. The peak containing Se-exDnaE was applied to a Superdex 75 column pre-equilibrated with 20 mM Tris–HCl (pH 7.5), 150 mM NaCl and eluted with

the same buffer. Further purification was carried out by applying the protein sample onto a Mono Q pre-packed column (Pharmacia) pre-equilibrated with 20 mM Tris–HCl (pH 7.5) and eluting with a linear gradient of NaCl.

The preDnaE and Se-exDnaE proteins were each concentrated to about 20 mg/ml before crystallization. Protein concentrations were estimated by measuring the absorbance at 280 nm.

Crystallization

Crystals of preDnaE were obtained by the hanging-drop, vapor-diffusion method at 18 °C by mixing 1 μl of protein solution containing 10 mM Tris (pH 7.5), 50 mM NaCl, 100 μM $ZnCl_2$ and 1 μl of reservoir solution containing 100 mM Hepes (pH 7.5), 25–30% (v/v) polyethylene glycol 400 and 200 mM $CaCl_2$. For crystallization of selenomethionyl exDnaE, the protein solution containing 10 mM Tris (pH 7.5), 50 mM NaCl was mixed with 100 mM Hepes (pH 7.5), 1.7–2.2 M $(NH_4)_2SO_4$ and 2% (v/v) PEG 400 in an equal ratio.

Data collection

X-ray diffraction data for Se-exDnaE were collected at 100 K using a MAC Science DIP2040 detector on beamline BL44XU of SPring8 (Hyogo, Japan). The Se-exDnaE crystal was transferred to a cryo-protectant solution containing 100 mM Hepes (pH 7.5) and 2.4 M $(NH_4)_2SO_4$ for data collection and diffracted to 1.67 Å. All intensity data were processed and scaled using the program HKL2000.[42] The X-ray diffraction data for preDnaE were collected at 100 K using an MAR CCD detector on beamlines 3W and 1A at BSRF (Beijing Synchrotron Radiation Facility, Beijing, China). The preDnaE crystal was flash-cryocooled for data collection and diffracted to 1.95 Å resolution. All intensity data were processed and scaled using the programs DENZO and SCALEPACK.[42] Data collection statistics are summarized in Table 2.

Structure determination and refinement

The structure of Se-exDnaE was solved initially by Se-single wavelength anomalous diffraction (SAD) method at 4 Å. Two selenium atoms were found in the asymmetric unit by the CNS heavy-atom search protocol and were used to phase the model.[43] Subsequent solvent flattening and density modification of the initial SAD phases yielded an experimental map for model building and refinement. There are two molecules in one asymmetric unit: the first was built manually, guided by the Ssp DnaB intein structure,[9] and the second was generated by molecular replacement in CNS.[43] Some (10%) of the data were set aside to calculate the free *R*-factor. The refinement was completed by alternating between manual building and minimization using data in the resolution range 50–1.67 Å. Group *B*-factor and individual *B*-factor refinement were then used to refine the temperature factors in the model. The final 1.67 Å refined structure of exDnaE consists of 159 amino acid residues and 242 water molecules. The quality of the final structure was verified with PROCHECK.[44] The Ramachandran plot shows that 96.7% of residues are in the most favorable regions and 3.3% in allowed regions.

The structure of preDnaE was solved by molecular replacement with CNS,[43] using exDnaE as a search model. There is one molecule per asymmetric unit. The final model of preDnaE, lacking residues -10, -9, 99-108 and 163-167, shows good stereochemistry with over 93.5% of the residues in the most favored region of the Ramachandran plot.

Table 2. Data collection and refinement statistics

Parameter	Se-exDnaE	preDnaE
A. *Data collection*		
Beamline	SPring8 BL44XU	BSRF 3W1A
Wavelength (Å)	0.97916	0.9806
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters		
a (Å)	50.2	42.5
b (Å)	67.0	58.2
c (Å)	90.3	67.1
$\alpha=\beta=\gamma$ (deg.)	90	90
Resolution (Å)[a]	50–1.67 (1.73–1.67)	50–1.95 (2.0–1.95)
Total observations	236,753	84,575
Unique reflections	35,350	12,680
Redundancy[a]	6.7 (4.5)	6.7 (6.2)
Matthews coefficient		
mol/asym	2	1
Matthews coefficient	2.1	2.1
Solvent (%, v/v)	41.7	40.9
Completeness[a](%)	97.9 (81.3)	99.9 (99.4)
R_{merge}[a,b]	5.3 (17.7)	6.4 (30.1)
Average $I/\sigma(I)$[a]	59.4 (10.1)	28.2 (6.2)
B. *Refinement*		
Resolution (Å)	50–1.67	50–1.95
R_{work}[c](%)	17.7	19.5
R_{free}[d](%)	20.2	22.2
Mean *B*-value (Å^2)		
All protein atoms	19.4	21.2
All atoms	20.7	21.5
Rmsd from ideality		
Bond lengths (Å)	0.020	0.020
Bond angles (deg.)	1.92	2.01

[a] Values in parentheses refer to the highest of the resolution bins.

[b] $R_{merge}=[\Sigma_{hkl}\Sigma_i|I_i(hkl)-\langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i\langle I(hkl)\rangle]\times100$, where $I_i(hkl)$ is the *i*th intensity measurement of reflection *hkl* and $\langle I(hkl)\rangle$ is the mean of the observations $I_i(hkl)$.

[c] $R_{work}=\Sigma||F_o|-|F_c||/\Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively.

[d] R_{free} was calculated based on 10% of the total data omitted during structure refinement.

Protein Data Bank accession codes

The coordinates and structure factors for preDnaE (code 1ZDE) and exDnaE (code 1ZD7) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University New Brunswick, NJ

Acknowledgements

We thank Dr Donald Comb and New England Biolabs, Inc. for generous support. We thank Wei Peng for his technical assistance. We are grateful to Fei Sun, Yujia Zhai and Dr Yamashita for their kind help with data collection and processing at SPring 8. We thank Dr Mark Bartlam, Dr Yiwei Liu and Dr Inca Ghosh for useful discussion. This research was supported by the NSFC (Grant No. 30221003 and

30128002), "863" Project (Grant No. 2002BA711A12) and "973" Project (Grant No. G1999075602).

References

1. Hirata, R., Ohsumk, Y., Nakano, A., Kawasaki, H., Suzuki, K. & Anraku, Y. (1990). Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **265**, 6726–6733.
2. Kane, P. M., Yamashiro, C. T., Wolczyk, D. F., Neff, N., Goebl, M. & Stevens, T. H. (1990). Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science*, **250**, 651–657.
3. Perler, F. B., Davis, E. O., Dean, G. E., Gimble, F. S., Jack, W. E., Neff, N. *et al*. (1994). Protein splicing elements: inteins and exteins – a definition of terms and recommended nomenclature. *Nucl. Acids Res.* **22**, 1125–1127.
4. Perler, F. B. (2002). InBase: the intein database. *Nucl. Acids Res.* **30**, 383–384.
5. Paulus, H. (2000). Protein splicing and related forms of protein autoprocessing. *Annu. Rev. Biochem.* **69**, 447–496.
6. Perler, F. B. (1998). Protein splicing of inteins and hedgehog autoproteolysis: structure, function, and evolution. *Cell*, **92**, 1–4.
7. Paulus, H. (2001). Inteins as enzymes. *Bioorg. Chem.* **29**, 119–129.
8. Perler, F. B. (2000). InBase, the Intein Database. *Nucl. Acids Res.* **28**, 344–345.
9. Ding, Y., Xu, M. Q., Ghosh, I., Chen, X., Ferrandon, S., Lesage, G. & Rao, Z. (2003). Crystal structure of a mini-intein reveals a conserved catalytic module involved in side-chain cyclization of asparagine during protein splicing. *J. Biol. Chem.* **278**, 39133–39142.
10. Ghosh, I., Sun, L. & Xu, M. Q. (2001). Zinc inhibition of protein trans-splicing and identification of regions essential for splicing and association of a split intein. *J. Biol. Chem.* **276**, 24051–24058.
11. Chong, S., Williams, K. S., Wotkowicz, C. & Xu, M. Q. (1998). Modulation of protein splicing of the *Saccharomyces cerevisiae* vacuolar membrane ATPase intein. *J. Biol. Chem.* **273**, 10567–10577.
12. Xu, M. Q. & Perler, F. B. (1996). The mechanism of protein splicing and its modulation by mutation. *EMBO J.* **15**, 5146–5153.
13. Mills, K. V., Dorval, D. M. & Lewandowski, K. T. (2005). Kinetic analysis of the individual steps of protein splicing for the *Pyrococcus abyssi* PolII intein. *J. Biol. Chem.* **280**, 2714–2720.
14. Duan, X., Gimble, F. S. & Quiocho, F. A. (1997). Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity. *Cell*, **89**, 555–564.
15. Hu, D., Crist, M., Duan, X., Quiocho, F. A. & Gimble, F. S. (2000). Probing the structure of the PI-SceI-DNA complex by affinity cleavage and affinity photocross-linking. *J. Biol. Chem.* **275**, 2705–2712.
16. Poland, B. W., Xu, M. Q. & Quiocho, F. A. (2000). Structural insights into the protein splicing mechanism of PI-SceI. *J. Biol. Chem.* **275**, 16408–16413.
17. Mizutani, R., Nogami, S., Kawasaki, M., Ohya, Y., Anraku, Y. & Satow, Y. (2002). Protein-splicing reaction via a thiazolidine intermediate: crystal structure of the VMA1-derived endonuclease bearing the N and C-terminal propeptides. *J. Mol. Biol.* **316**, 919–929.
18. Moure, C. M., Gimble, F. S. & Quiocho, F. A. (2002). Crystal structure of the intein homing endonuclease PI-SceI bound to its recognition sequence. *Nature Struct. Biol.* **9**, 764–770.
19. Werner, E., Wende, W., Pingoud, A. & Heinemann, U. (2002). High resolution crystal structure of domain I of the *Saccharomyces cerevisiae* homing endonuclease PI-SceI. *Nucl. Acids Res.* **30**, 3962–3971.
20. Ichiyanagi, K., Ishino, Y., Ariyoshi, M., Komori, K. & Morikawa, K. (2000). Crystal structure of an archaeal intein-encoded homing endonuclease PI-PfuI. *J. Mol. Biol.* **300**, 889–901.
21. Klabunde, T., Sharma, S., Telenti, A., Jacobs, W. R., Jr & Sacchettini, J. C. (1998). Crystal structure of GyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing. *Nature Struct. Biol.* **5**, 31–36.
22. Hall, T. M., Porter, J. A., Young, K. E., Koonin, E. V., Beachy, P. A. & Leahy, D. J. (1997). Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins. *Cell*, **91**, 85–97.
23. Wu, H., Hu, Z. & Liu, X. Q. (1998). Protein trans-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. *Proc. Natl Acad. Sci. USA*, **95**, 9226–9231.
24. Evans, T. C., Jr, Martin, D., Kolly, R., Panne, D., Sun, L., Ghosh, I. *et al*. (2000). Protein trans-splicing and cyclization by a naturally split intein from the dnaE gene of *Synechocystis* species PCC6803. *J. Biol. Chem.* **275**, 9091–9094.
25. Mills, K. V. & Paulus, H. (2001). Reversible inhibition of protein splicing by zinc ion. *J. Biol. Chem.* **276**, 10832–10838.
26. Nichols, N. M., Benner, J. S., Martin, D. D. & Evans, T. C., Jr (2003). Zinc ion effects on individual Ssp DnaE intein splicing steps: regulating pathway progression. *Biochemistry*, **42**, 5301–5311.
27. Nichols, N. M. & Evans, T. C., Jr (2004). Mutational analysis of protein splicing, cleavage, and self-association reactions mediated by the naturally split Ssp DnaE intein. *Biochemistry*, **43**, 10265–10276.
28. Martin, D. D., Xu, M. Q. & Evans, T. C., Jr (2001). Characterization of a naturally occurring trans-splicing intein from *Synechocystis* sp. PCC6803. *Biochemistry*, **40**, 1393–1402.
29. Mizutani, R., Anraku, Y. & Satow, Y. (2004). Protein splicing of yeast VMA1-derived endonuclease via thiazolidine intermediates. *J. Synchrotron Radiat.* **11**, 109–112.
30. Romanelli, A., Shekhtman, A., Cowburn, D. & Muir, T. W. (2004). Semisynthesis of a segmental isotopically labeled protein splicing precursor: NMR evidence for an unusual peptide bond at the N-extein-intein junction. *Proc. Natl Acad. Sci. USA*, **101**, 6397–6402.
31. Gogarten, J. P., Senejani, A. G., Zhaxybayeva, O., Olendzenski, L. & Hilario, E. (2002). Inteins: structure, function, and evolution. *Annu. Rev. Microbiol.* **56**, 263–287.
32. Chen, L., Benner, J. & Perler, F. B. (2000). Protein splicing in the absence of an intein penultimate histidine. *J. Biol. Chem.* **275**, 20431–20435.
33. Caspi, J., Amitai, G., Belenkiy, O. & Pietrokovski, S. (2003). Distribution of split DnaE inteins in cyanobacteria. *Mol. Microbiol.* **50**, 1569–1577.

34. Alberts, I. L., Nadassy, K. & Wodak, S. J. (1998). Analysis of zinc binding sites in protein crystal structures. *Protein Sci.* **7**, 1700–1716.
35. Shingledecker, K., Jiang, S. & Paulus, H. (2000). Reactivity of the cysteine residues in the protein splicing active center of the *Mycobacterium tuberculosis* RecA intein. *Arch. Biochem. Biophys.* **375**, 138–144.
36. Sun, W., Yang, J. & Liu, X. Q. (2004). Synthetic two-piece and three-piece split inteins for protein trans-splicing. *J. Biol. Chem.* **279**, 35281–35286.
37. Scott, C. P., Abel-Santos, E., Wall, M., Wahnon, D. C. & Benkovic, S. J. (1999). Production of cyclic peptides and proteins *in vivo*. *Proc. Natl Acad. Sci. USA*, **96**, 13638–13643.
38. Chen, L., Pradhan, S. & Evans, T. C., Jr (2001). Herbicide resistance from a divided EPSPS protein: the split Synechocystis DnaE intein as an *in vivo* affinity domain. *Gene*, **263**, 39–48.
39. Chin, H. G., Kim, G. D., Marin, I., Mersha, F., Evans, T. C., Jr, Chen, L. *et al*. (2003). Protein trans-splicing in transgenic plant chloroplast: reconstruction of herbicide resistance from split genes. *Proc. Natl Acad. Sci. USA*, **100**, 4510–4515.
40. Evans, T. C., Jr & Xu, M. Q. (1999). Intein-mediated protein ligation: harnessing nature's escape artists. *Biopolymers*, **51**, 333–342.
41. Xu, M. Q., Paulus, H. & Chong, S. (2000). Fusions to self-splicing inteins for protein purification. *Methods Enzymol.* **326**, 376–418.
42. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Macromol. Crystallog. A*, **276**, 307–326.
43. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al*. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
44. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.

Edited by R. Huber

(Received 26 June 2005; received in revised form 21 August 2005; accepted 13 September 2005)
Available online 30 September 2005